COMMENTS ON “ANALYSIS OF TRADITIONAL AND REVERBERATION-REDUCING METHODS OF ROOM EQUALIZATION” [1]

I have read with interest the above paper and have found that the topics discussed provide useful insight in many aspects of digital room equalization. However, I have also found that the paper has overlooked previously-published work in this area, so that in my opinion, some of the conclusions drawn are mainly relevant to the specific methodology adopted by the paper’s author and thus are not generally applicable to the problem of reverberation-reducing methods.

Aim of this letter, is to amend this shortcoming, to explain that some of the findings in the paper were known for a very long time, and also to illustrate that appropriate solutions have been proposed which alleviate the problems discussed in the above paper.

(a) A main conclusion derived in the paper is the detrimental effect of “high-Q” Room Transfer Function (RTF) zeros on the derivation of the inverse filter, due to the duration of the compensating “ringing” poles. Although this is a correct and useful conclusion, it is by no means novel, since it is well-known that the inverse for any such finite length function (as is a measured room impulse response) can be potentially of infinite duration (see [2], p.209, [8] p.896). The paper’s author has opted for a DFT-based method for obtaining such inverses ([1], Section 2, eq. (4)), which as will be discussed here, may lead to misleading conclusions.

As early as 1982 [3], but also in a later JAES paper [4], this author and his co-workers have indicated that such an inversion methodology is potentially improper for practical, mixed-phase responses, since, as was stated in [4]: “…the inverse operator will usually be infinite in length, even if the original response is finite. Consequently implementation using finite-length discrete Fourier transforms leads to truncation of higher order terms, which is manifested in the inverse operator as a form of aliasing…”. To solve this problem, we had proposed a least-squares inversion technique, based on the “Simpson sideways recursion” extension of the more well-known Levinson recursion [3, 4], which evaluates an optimal finite length inverse, rendered causal with the introduction of an appropriate delay function to account for the “pre-ringing” associated with the acausal component of the room response inverse.

As a result of this, inverses derived via the least-squares approach are well-behaved both in terms of containing the effects of ringing poles which are generated to compensate for “high-Q zeros”,
since the inversion performance is constrained by the chosen filter length, N so that such extremely “sharp” zeros can be only partially compensated, and also in terms of the choice of the delay, which in most practical cases is not symmetrically located for the causal and acausal component of a room response. Nevertheless, as was reported in [5], by using a least-squares derived inverse, the dereverberated signal was still audibly distorted.

The paper’s author instead, has adopted a DFT-based inversion, rendered symmetric for the causal and acausal component and constrained by the use of “regularization” method. Given that “regularization” is a non-linear operation on the spectrum with often unpredictable artifacts, it is important here to concentrate for the case in [1] when $\beta= -80$ dB, i.e. when according to the author ([1], page 11) and also according to Figure 10, there is virtually no regularization effect on the DFT-based inversion result, which is shown in Figure 12 (a). As can be observed in this figure, there is no visible tapering of the inverse’s energy at the filter’s causal and acausal time edges, leading to the assumption that such “ideal” inversion may create aliasing / truncation components due to the fact that the potentially infinite in length inverse function is derived via the inverse transformation of a finite-length DFT. It could be argued that in practice such effects would be of little importance, given the extreme size of the DFTs employed by the author, but it appears here that as a result of this characteristic of the inverse filter, some side-effects are generated after its deconvolution with the original response. Specifically, convolution with of this inverse with a measured response to derive the time domain responses after dereverberation, generates peaks at the edges of the response, in my opinion related to the above truncation effect. Such peaks are manifested largely in Figure 14(a) and to a lesser extend in Figures 14 (b) and (c), where such inversion aliasing / truncation effects are reduced, by the expected reduction of inverse length.

Hence, in my opinion, such boundary artifacts (peaks) are not a genuine feature of dereverberation methods, but a by-product of the inversion implementation method chosen by the author.

For comparison, I present below, inversion results of a measured room impulse response, corresponding to a professional listening room of dimensions very similar to those of the room used in [1], i.e. L: 7,15m X W: 4,60m X H: 2,90m and acoustic properties comparable to those employed in the paper [1] (see Table 1, below). The processing methodology and parameters were identical to those used by the author (i.e. 10,560 sample point room response, 513,729 sample point inverse filter). Results for two cases are presented: (i) DFT-based inversion with no
regularization (in practice similar to the case $\beta=-80$ dB, in the paper [1]), (ii) least-squares based inversion [3].

Figure 1(a) presents the time-domain response for the DFT-based inversion and Figure 1(b) the time-domain response for the least squares-based inversion. It is useful to observe that the DFT-based inverse exhibits the previously discussed potentially aliased behavior due to the truncation of the DFT products (being similar to the Figure 12(a)), whereas the Figure 1(b) indicates the previously explained time-constrained behavior at both time edges.

After convolution with the measured response (in a way identical to the description in Section 4.2.4 of the paper [1]), the time-domain response after dereverberation is derived for the two alternative inversion methodologies, shown respectively in Figures 2(a) and 2(b). As it can be observed in Figure 2(a), boundary artifacts are present, similar to those shown by the author in Figure 14(a), whereas such peaks are absent for the case of least-squares inversion (Figure 2(b)). Nevertheless, use of these filters for real-time reproduction of pre-processed audio, generates perceptually-detrimental artifacts, so that such distortions cannot be related to the boundary peaks (which, if present, are expected to contribute additional audible effects).

(b) The above discussion introduces questions on the analysis of the perceptually-detrimental dereverberation artifacts, as is presented in pages 14 and 15 of the paper [1]. In my opinion, such perceptually-detrimental artifacts will exist in all “ideal” dereverberation methods (as I had also found in 1985, [5]), but for reasons different to those explained by the author (i.e. the existence of boundary artifacts, as was discussed above), since even if alternative inversion methodologies for such “ideal” dereverberation case were adopted, when such artifacts would not be present, strong audible coloration and smearing would still appear. I believe that the reasons for such distortions, which incidentally do not affect anechoically measured and equalized loudspeaker responses, have yet to be properly explained.

Apart from this point, Section 3 of the paper offers a valuable, if rather simplified methodology for evaluation of the subjective results of dereverberation. In addition to such a simple model, I want to draw the attention to a more comprehensive room masking model, introduced in [6] and not examined in the paper [1]. Still, even if this more detailed model produces results which fit well with many published psychoacoustic data on reflection masking, I feel that the currently known
psychoacoustic principles cannot be employed to accept or reject the potential uses of dereverberation, since at this stage in their evolution they can be only used as a rough guide and they cannot fully explain these known audible dereverberation distortions. Clearly, more research work is required in this direction.

(c) Such problems with “ideal” inverse dereverberation, have forced this author, my co-workers along with many other researchers, over the subsequent years, to examine and publish alternative inversion strategies. A limited overview of such methods is presented and discussed by the author ([1], page 6). However, many other reverberation-reducing methods have been overlooked. For example, pole/zero response modeling and inversion was proposed by us in 1991 [7], codebook-based multipoint room equalization of low-order minimum-phase room response in 1992 [8], target-constrained response inversion in 1995 [9], and complex smoothing for the modification of measured room responses in 1999 [10], with recent results for use of such responses for dereverberation [11]. Miyoshi and Kaneda [18] have considered a method with inverse derived from multiple FIR filters and additional signal reproducing channels. Multiband response inversion was discussed in [19], warped-frequency scale deconvolution was presented in [20], and recently, low-frequency modal equalization in [21]. Equalization of car acoustics was extensively studied by Farina and his co-workers (e.g. [22]), whereas, recently [23], a psychoacoustically-optimized equalization method has been also proposed. Possibly, other proposed methods have been also presented, but are unintentionally missed from this short list.

(d) With respect to Section 4.3 of the paper [1] (Effect of Physical Displacement), again, I believe that a significant volume of previous work on this aspect of dereverberation has been overlooked by the paper’s author, so that at one hand the findings presented in the paper are not novel, and at the other hand proposed solutions for the described problems by other workers, have not been discussed in the paper.

Approximately 18 years ago [5], this author has studied - with far less suitable technical means - the effect of source / receiver mismatch in the dereverberation. My findings were very similar to those reported in the paper: “…however the processed signal was not completely free from reverberation; significantly, frequency domain distortions (coloration) were not removed, and in fact, they were often reinforced after processing….in order to obtain a consistent improvement from inversion, the response used must be measured no further apart that ½ critical distance units
Clearly, such condition cannot be satisfied in many practical situations, indicating the limited scope for reverberant signal enhancement by response deconvolution methods…”. More recently, Radlovic et. al. [12], (reference [33] in [1]), have confirmed those findings, so that in this respect, the paper provides little novel evidence for this problem.

In the intervening period, some proposals have been also made for alleviating the problem of source / receiver displacements, which are not discussed in the paper. This author has proposed a codebook-based solution for low-order magnitude Room Transfer Function (RTF) classification [8], Wilson has proposed a response averaging approach mainly for off-axis loudspeaker response compensation [13], a common-pole RTF inversion approach was proposed in [14], Asano et. al. [15] have proposed a equalization approach based on derivative constraints and lately, Bhariktar and Kyriakakis [16] have considered equalization based on fuzzy-clustering. It also appears [11], that as long as “non-ideal” inversion is required, displacement sensitivity is not an issue as serious as the early studies and the above paper [1] suggests, but at one hand the mismatch can be moderated by the response simplification pre-processing (see discussion below) and at the other hand, it may be addressed by suitable algorithms, as proposed in the above references.

(e) By overlooking many relevant references, as was shown, the paper is not making evident to the reader that a significant trend for contemporary dereverberation methods, especially during the last few years, is for inverting appropriately-modified versions of measured loudspeaker-room responses. These methods usually avoid compensating for the problematic “high-Q” response zeros, which incidentally are manifested more dramatically in high-resolution RTFs (i.e. derived from long DFTs, as was the case in [1]), and furthermore it is known to be of smaller perceptual significance (see reference [22] in [1]). This is achieved by either modifying measured responses [10, 11], or by introducing alternative measurement strategies [23], or by low-order modelling of RTFs [21, 7], or by appropriate time-frequency selective deconvolution (e.g. the work of Johansen and Rubak, i.e. references [16, 17] in the paper [1]). These methods achieve by definition a non-ideal RTF inversion and hence a smaller degree of enhancement than theoretically could be achieved by the “ideal” dereverberation case, without ‘though introducing, as the paper [1] concludes: “…extremely audible and annoying resonances…”, or even any other audible degradation.
We have recently produced results for such a method, based on the Complex Smoothing of room responses [10], and tested it over a large variety of different spaces [11]. Audio demonstrations were also provided in the recent AES Convention, but furthermore, such audio demos can be also found in the electronic address [17]. This method is free of all the problems (i.e. audible artifacts, displacement sensitivity, measurement variations) discussed in the paper [1]. To alleviate any potential scepticism with respect to this claim, I propose to the paper’s author to provide me with the response measurement(s) corresponding to the room(s) referred to in the paper. Then, it is promised that I shall return to him, pre-filtered and unfiltered audio files (of his choice) for real-time reproduction and evaluation within these room(s).

I want to conclude with the observation that for the benefit of past, present and future contributors in this field, the above points and omissions should have been considered and potentially corrected during the paper’s reviewing process.

REFERENCES


Table 1. Reverberation time versus frequency for the listening room

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Reverberation Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>0,468</td>
</tr>
<tr>
<td>250</td>
<td>0,396</td>
</tr>
<tr>
<td>500</td>
<td>0,349</td>
</tr>
<tr>
<td>1000</td>
<td>0,342</td>
</tr>
<tr>
<td>2000</td>
<td>0,424</td>
</tr>
<tr>
<td>4000</td>
<td>0,371</td>
</tr>
<tr>
<td>8000</td>
<td>0,292</td>
</tr>
<tr>
<td>average</td>
<td>0,368</td>
</tr>
</tbody>
</table>
Mourjopoulos J.: Comments on “Analysis of Traditional and Reverberation-Reducing Methods of Room Equalization”

Figure 1: Time-domain responses of inversion filters of a professional listening room, with alternative inversion methods. (a) DFT inversion. (b) least-squares inversion
Figure 2: Time-domain responses for professional listening room after dereverberation with the alternative inverse filters. (a) deconvolution with the DFT inverse. (b) deconvolution with the least-squares inverse.

John N. Mourjopoulos, AES Member
Audio Group
Wire Communications Laboratory
Electrical and Computer Engineering Department
University of Patras
Patras 26500, Greece